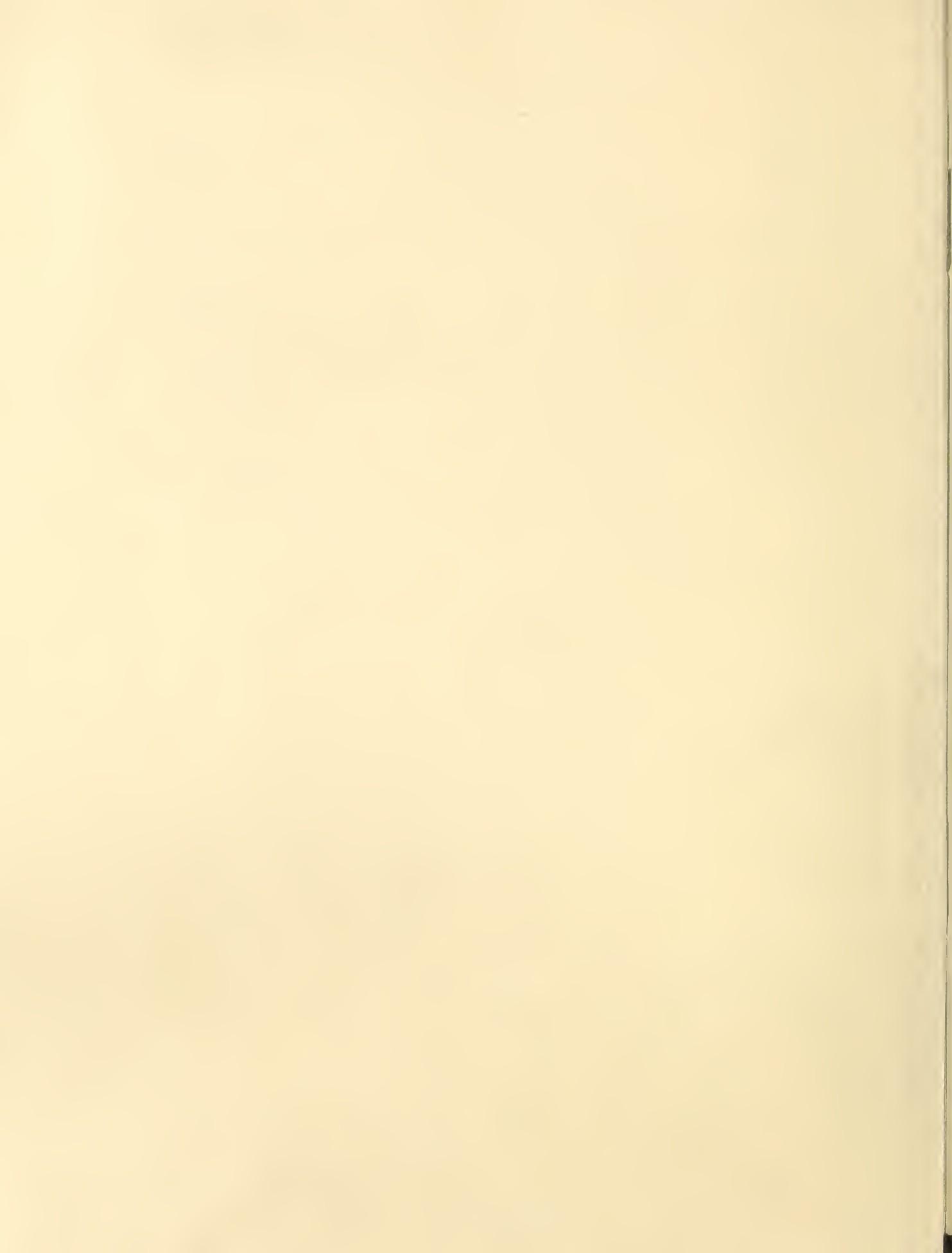


Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.



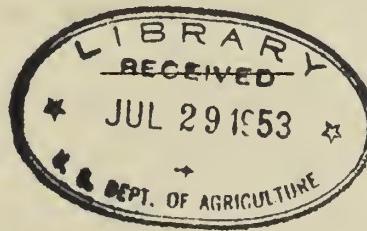
File #

NOT FOR GENERAL DISTRIBUTION

UNITED STATES DEPARTMENT OF AGRICULTURE
Agricultural Marketing Service

A SURVEY OF EXPERIMENTAL DESIGN

By
W. G. Cochran,
Professor of Mathematical Statistics,
Iowa State College



Washington, D. C.
March 1940

Prefatory Note

The material in the following pages was presented to research workers of the Department in a series of six lectures as part of an in-service training program in the design of experiments.

The lectures were arranged under the joint auspices of the Graduate School of the U. S. Department of Agriculture and the Project on Statistical Methods, which is a cooperative undertaking carried on by the Agricultural Marketing Service of the Department and the Statistical Laboratory of Iowa State College.

The subject of experimental design is of considerable interest at the present time and one which Mr. Cochran is exceptionally well qualified to discuss.

Charles F. Sarle,
Commodity Credit Corporation,
U. S. Department of Agriculture.

S16780

A SURVEY OF EXPERIMENTAL DESIGN

By W. G. Cochran, Professor of Mathematical Statistics,
Iowa State College 1/

These notes are taken from a series of six lectures given in the U. S. Department of Agriculture, Washington, in January 1940. They are intended for students who have taken an elementary course in experimental design. Their aim is to serve as a guide to research workers in deciding what type of design is most likely to suit the needs of a particular experiment. The methods of analysing the results are not discussed except insofar as is necessary in explaining the properties of the designs; they will be found in the references given at the end. One or two notes which were not given in the lectures have been added for the sake of completeness.

All the designs to be described below are based on two standard types-- Randomized blocks and the Latin square.

Randomized Blocks

The site of the experiment is divided into a number of compact blocks, each block containing as many plots as there are treatments. Treatments are assigned at random to the plots in each block. The advantages of this type of design may be classed under three headings:

1. Accuracy. The device of dividing the experimental material into groups or blocks offers the prospect of increasing the accuracy of comparisons between treatments, since differences between blocks are eliminated from the sources of experimental error.

2. Flexibility. The design places no restriction on the number of treatments or on the number of replications. In general, however, at least two replications are required to obtain tests of significance.

3. Ease of analysis. The statistical analysis is simple and rapid. Moreover, the error of any treatment comparison can be isolated, and any number of treatments may be omitted from the analysis without complicating it. These facilities may be useful when certain treatment differences turn out to be very large, when some treatments produce crop failures, or when the experimental material is heterogeneous.

1/. Professor Cochran was formerly assistant statistician, Rothamsted Experimental Station, Harpenden, England.

Amount of Replication

The amount of replication which is advisable depends on several factors--cost, labor, variability of the material, probable size of the treatment differences, and the standard of accuracy aimed at. As regards the purely statistical side of the question, some idea may be gained of the size of treatment differences which will be detected as significant, if the probable size of the standard error per plot is approximately known from previous experiments on similar material. For example, if the standard error per plot is likely to be about 8 percent and 6 replications are to be used, the standard error of a treatment mean is $8/\sqrt{6}$ percent, or 3.3 percent. Thus an observed difference of about $3 \times 3.3\frac{2}{3}$, or 10 percent, between two treatment means will be significant at the 5 percent level. A simple calculation of this type gives some idea of the discriminating power of the experiment, and may be useful in avoiding experiments which from the start have little chance of detecting small treatment effects.

The above calculation does not mean that a true difference of 10 percent between two treatments is certain to be detected. If the true difference is 10 percent, the observed difference may be either above or below it, owing to experimental errors. If these are symmetrically distributed, the chance of detecting a true difference of 10 percent is 1 in 2 in a single experiment. By an extension of the above calculation, the chance of detecting a true difference of any given magnitude can be found. This calculation is occasionally helpful in settling the amount of replication in crucial experiments.

Shape of Block and of Plots Within the Block

Blocks should be placed so as to make the differences between them as large as possible. Thus if an experiment is to be conducted on a hillside, and the fertility of the soil is likely to change as we ascend the slope, plots in the same block would be placed at the same distance up the slope; i.e., the block would lie perpendicular to the slope..

On a level field, if appearance and previous history give no knowledge about the fertility gradients, each block is made as compact as possible, square blocks being advisable if they can be fitted into the experimental site.

The object in deciding the shape of plots within the block is to make all plots in the same block as alike as possible in yield. For this reason, plots which extend the whole length of one side of the block, as in figure I (a), are preferred to plots which are compact, as in figure I (b).

2/. For experiments with less than 12 degrees of freedom for error, the quantity " $\sqrt{2} \times 5$ percent point of t " should be used instead of 3 in multiplying the standard error.

A SURVEY OF EXPERIMENTAL DESIGN

By W. G. Cochran, Professor of Mathematical Statistics,
Iowa State College 1/

These notes are taken from a series of six lectures given in the U. S. Department of Agriculture, Washington, in January 1940. They are intended for students who have taken an elementary course in experimental design. Their aim is to serve as a guide to research workers in deciding what type of design is most likely to suit the needs of a particular experiment. The methods of analysing the results are not discussed except insofar as is necessary in explaining the properties of the designs; they will be found in the references given at the end. One or two notes which were not given in the lectures have been added for the sake of completeness.

All the designs to be described below are based on two standard types-- Randomized blocks and the Latin square.

Randomized Blocks

The site of the experiment is divided into a number of compact blocks, each block containing as many plots as there are treatments. Treatments are assigned at random to the plots in each block. The advantages of this type of design may be classed under three headings:

1. Accuracy. The device of dividing the experimental material into groups or blocks offers the prospect of increasing the accuracy of comparisons between treatments, since differences between blocks are eliminated from the sources of experimental error.

2. Flexibility. The design places no restriction on the number of treatments or on the number of replications. In general, however, at least two replications are required to obtain tests of significance.

3. Ease of analysis. The statistical analysis is simple and rapid. Moreover, the error of any treatment comparison can be isolated, and any number of treatments may be omitted from the analysis without complicating it. These facilities may be useful when certain treatment differences turn out to be very large, when some treatments produce crop failures, or when the experimental material is heterogeneous.

1/. Professor Cochran was formerly assistant statistician, Rothamsted Experimental Station, Harpenden, England.

Amount of Replication

The amount of replication which is advisable depends on several factors--cost, labor, variability of the material, probable size of the treatment differences, and the standard of accuracy aimed at. As regards the purely statistical side of the question, some idea may be gained of the size of treatment differences which will be detected as significant, if the probable size of the standard error per plot is approximately known from previous experiments on similar material. For example, if the standard error per plot is likely to be about 8 percent and 6 replications are to be used, the standard error of a treatment mean is $8/\sqrt{6}$ percent, or 3.3 percent. Thus an observed difference of about $3 \times 3.3\%$, or 10 percent, between two treatment means will be significant at the 5 percent level. A simple calculation of this type gives some idea of the discriminating power of the experiment, and may be useful in avoiding experiments which from the start have little chance of detecting small treatment effects.

The above calculation does not mean that a true difference of 10 percent between two treatments is certain to be detected. If the true difference is 10 percent, the observed difference may be either above or below it, owing to experimental errors. If these are symmetrically distributed, the chance of detecting a true difference of 10 percent is 1 in 2 in a single experiment. By an extension of the above calculation, the chance of detecting a true difference of any given magnitude can be found. This calculation is occasionally helpful in settling the amount of replication in crucial experiments.

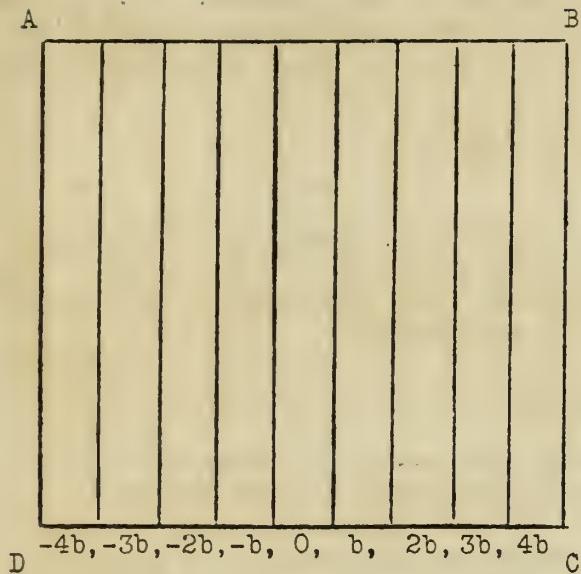
Shape of Block and of Plots Within the Block

Blocks should be placed so as to make the differences between them as large as possible. Thus if an experiment is to be conducted on a hillside, and the fertility of the soil is likely to change as we ascend the slope, plots in the same block would be placed at the same distance up the slope; i.e., the block would lie perpendicular to the slope.

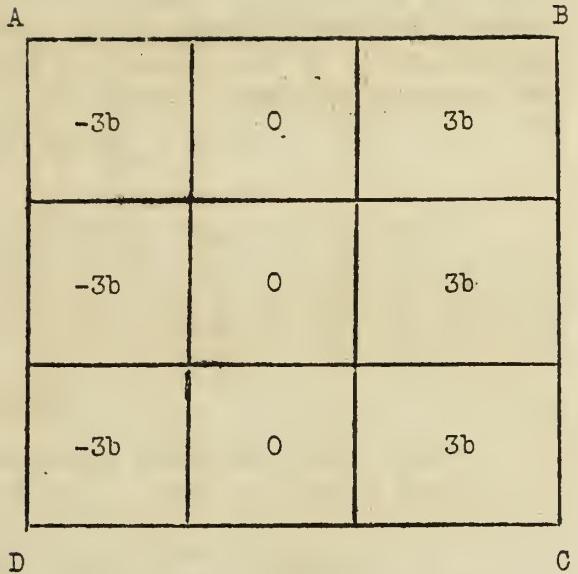
On a level field, if appearance and previous history give no knowledge about the fertility gradients, each block is made as compact as possible, square blocks being advisable if they can be fitted into the experimental site.

The object in deciding the shape of plots within the block is to make all plots in the same block as alike as possible in yield. For this reason, plots which extend the whole length of one side of the block, as in figure I (a), are preferred to plots which are compact, as in figure I (b).

2/. For experiments with less than 12 degrees of freedom for error, the quantity " $\sqrt{2} \times 5$ percent point of t" should be used instead of 3 in multiplying the standard error.



(a)



D

(b)

Figure I

The relative accuracy of the two shapes of plot in figure I may be examined theoretically. The type of soil variation which makes the two shapes differ most is a linear fertility gradient running parallel either to AB or AD. Suppose that there is such a gradient, and that the difference in yield between two neighboring plots of shape I(a) is b . If the gradient runs parallel to AD, it does not affect the differences between plots in I(a). If, however, the gradient runs parallel to AB, the mean yields of the nine plots may be written as shown in figure I(b). In this case the sum of squares within the block due to the gradient is

$$b^2 \left\{ (-4)^2 + (-3)^2 + \dots + (4)^2 \right\} = 60b^2, \text{ and the mean square is } \frac{60}{8}b^2.$$

In figure I(b), it does not matter whether the prevailing gradient is guessed correctly or incorrectly, since the plots are symmetrically placed with regard to both gradients. In either case, the sum of squares is

$$b^2 \left\{ (-3)^2 + (-3)^2 + (-3)^2 + (3)^2 + (3)^2 + (3)^2 \right\} = 54b^2.$$

If σ_r^2 is the random variance within blocks, independent of the shape of plot, the two arrangements give the following mean squares within the block.

Fertility gradient	I(a)	I(b)
parallel to AD	σ_r^2	$\sigma_r^2 + \frac{27}{4}b^2$
parallel to AB	$\sigma_r^2 + \frac{30}{4}b^2$	$\sigma_r^2 + \frac{27}{4}b^2$

If the gradient is marked, arrangement I(a) is definitely superior when the long side of the plot is parallel to the gradient, and little worse when the long side of the plot is perpendicular to the gradient. If, in the absence of prior knowledge, the plots are equally likely to be parallel or perpendicular to the prevailing gradient, the average error variances of the two arrangements may be taken as

$$\sigma_r^2 + \frac{15}{4} b^2 \quad \text{and} \quad \sigma_r^2 + \frac{27}{4} b^2.$$

Besides the above statistical considerations, questions of practicability must also be borne in mind in reaching a decision on the shape of plot and the arrangement of the plots in the block. For some crops a long narrow plot of the type required in I(a) would be unsuitable.

Latin Squares

The second standard type--the Latin square--may be compared with randomized blocks on each of the three criteria given for the former.

1. Accuracy. The Latin square offers the possibility of greater accuracy than randomized blocks, since it eliminates the effect of fertility variations in two directions on the treatment means.

2. Flexibility. Since the number of treatments is equal to the number of replications, this design is much less flexible than randomized blocks; squares of size larger than 10×10 are seldom used, on account of the high amount of replication which they entail.

3. Ease of analysis. The statistical analysis is easy to perform. However, the error term cannot be subdivided to give the particular error of a single treatment comparison, and the calculations when one or more treatments have to be omitted from the analysis are somewhat more involved than for randomized blocks.

Randomized blocks and Latin squares are generally considered most suitable for experiments where the number of treatments does not exceed twelve. With higher numbers, the Latin square is ruled out because of excessive replication, while randomized blocks tend to become less accurate because the plots in the same block can no longer be kept all near one another. Since modern experiments often involve large numbers of treatments, it is important to know by how much the standard error per plot is likely to increase as the number of plots in the block increases. It is equally important to study the relative accuracy of randomized blocks and the Latin square, since many experiments could be put in either form of design. Unfortunately, the available information on these questions is scanty. It may be instructive to consider how these questions could be examined. There are two methods.

(a). From uniformity-trial data. Comparable Latin square and randomized block designs could be superimposed on the results of uniformity trials. Similarly, the variation within blocks containing different numbers of plots could be calculated. The limitations to the use of this approach are that the amount of uniformity-trial data on any one crop is rather meager, and also that there has been a tendency, in laying out uniformity-trials, to select apparently uniform fields, so that the results may not be representative of normal experimental conditions.

(b). From field experiments. Every replicated experiment of the randomized blocks or Latin square type gives some information on the question of block size. Suppose that a Randomized blocks experiment with 4 blocks of 6 treatments each gives the following analysis of variance:

	Degrees of freedom	Mean squares
Blocks	3	100
Treatments	5	145
Error	15	20

We observe that the blocks have been fairly successful in removing soil heterogeneity. We can go further and estimate what the error mean square might have been had the experiment been completely randomized within the 24 plots, instead of dividing it into six blocks. In this case, the degrees of freedom between blocks would not be taken out of error, and the answer at first sight might appear to be:

$$\frac{(3 \times 100) + (15 \times 20)}{18} = 33.3.$$

This error is the total variation within treatments, as it should be for a completely randomized experiment; it was, however, calculated from an experiment in which each treatment occurred once in each block. Thus it is an appropriate error only for those completely randomized designs which happen to turn out as randomized blocks.

To calculate an average error for all completely randomized designs, we may note that in these designs no attempt is made to associate any particular degrees of freedom with treatments. (In randomized blocks, on the other hand, we insist that the treatment degrees of freedom shall come entirely from within blocks.) Thus any set of degrees of freedom, such as blocks, will contribute to the average error in proportion to the number of degrees of freedom which it represents. The exact contribution of the 5 treatment degrees of freedom in this experiment is unknown, but as these are "within block" comparisons, the best estimate of their error mean square is obtained from the 15 degrees of freedom for error in the experiment. Hence the average error is estimated as:

$$\frac{(3 \times 100) + (20 \times 20)}{23} = 30.$$

In this experiment, two replications in blocks of six plots would have been as accurate as three in blocks of twenty-four plots--a handsome increase in accuracy for the reduction of block size.

Similarly, any Latin square can be compared with randomized blocks parallel either to the rows or to the columns, and with complete randomization within the whole site of the experiment. These comparisons have the advantage of being made from actual experiments, though the type of comparison supplied is decided by the experiment, and is outside our control. An experiment station which makes routine calculations of this type will amass a large and useful body of miscellaneous information, with little extra labor.

For the field experiments carried out at Rothamsted and associated centers between 1927 and 1934, the following are comparable error mean squares for the three types:

Completely randomized	Randomized blocks	Latin squares
100	60	45

Ten replications of a completely randomized design were about as accurate as six in randomized blocks and four or five in Latin squares. The figures also indicate a definite superiority of the Latin square over randomized blocks, though I do not have data to show whether this is typical of conditions in this country. If four is taken as an average number of replications in these trials, the randomized blocks figures suggest that the reduction of block size to one-quarter reduces the variance by 40 percent.

The Split-Plot Design

In this there are two groups of treatments: the whole, or main-plot, and the sub-plot treatments. The former are applied to whole plots, which may be arranged either in randomized blocks or in a Latin square. Each whole plot is divided into a number of sub-plots equal to the number of sub-plot treatments, and the latter are allocated at random to these sub-plots. In the analysis of variance, two errors are obtained, one applicable to differences between whole-plot treatments, and the other to differences between sub-plot treatments and to the interactions between main- and sub-plot treatments. Since the sub-plot error arises only from differences between sub-plots in the same main-plot, it is usually smaller than the main-plot error. The consequence is that sub-plot treatments and the interactions between the two groups are compared more precisely than main-plot treatments. The principal considerations affecting the utility of these designs may be classed under three heads.

1. Accuracy. This design is clearly appropriate where one group of treatment comparisons is not required with any great precision, the principal aim being to test the second group and its interactions with the first. This might be the case if the main-plot treatment was a fertilizer, say phosphate, of known efficacy, which was fairly certain to be applied in practice. The object of the experiment might be to test whether there was any response to another treatment, such as lime, and whether the response varied over the range of dressings of phosphate likely to be used in practice. In this case the main-plot treatments might consist of increasing levels of phosphate, while the lime treatments would be applied to sub-plots.

2. Convenience. Sometimes a particular type of treatment cannot satisfactorily or conveniently be applied on very small plots. Examples are cereals which are sown with a drill, cultivation implements which require a turning headland, and manure, which is difficult to weigh and spread evenly over small plots. Such treatments may conveniently be applied to main-plots, provided that sufficient replication is carried to secure a reasonable degree of accuracy.

3. Possible use of Latin squares. Suppose that one group of treatments contains five treatments, and the other three, and that all combinations of the two groups are to be included, making 15 treatments in all. This could be laid out in randomized blocks of 15 plots each, but this design might not be very accurate, owing to the rather large number of plots in the block. Moreover, no reduction of block size by confounding is possible without confounding main effects. If, however, five replications can be made, the first group of treatments might be put on main plots in a 5 x 5 Latin square, and the second group on sub-plots.

It may be objected that this sacrifices accuracy on the main effects of the group of five treatments. This is not necessarily so, because the use of the Latin square, instead of randomized blocks may compensate for the possible loss of precision. Some information may be gained on this point from a study made by Yates. From the results of a number of split-plot Latin squares, he estimated the error variances which would have been obtained if both main- and sub-plot treatments had been put in the same randomized block design. If the error variance of the randomized block design is taken as 100, the error variances of the main- and sub-plot treatments are shown below for two sets of split-plot Latin square experiments.

Plots split into	Main Plots	Error variance of Sub-Plots
Two	(22) 80	60
Four	(9) 100	80

The figures in brackets are the number of experiments included. Thus, for experiments in which the plot was split into two, both main- and sub-plot treatments were substantially more accurately compared than they would have been if combined in the same randomized block design. For splits into four, the main-plot treatments were still no worse off as compared with randomized blocks, while the sub-plot treatments, as one would expect, were somewhat more accurately determined. These results also substantiate the superiority of the Latin square over randomized blocks, at least as far as the Rothamsted data are concerned.

Factorial Design

Where experiments are to be carried out on the effects of a number of different factors, much saving of expense and labor may result by testing the different factors, in all combinations, in the same experiment. For instance, suppose that the responses to given dressings of nitrogen, potash, and phosphate are to be tested. If all eight combinations of the absence and presence of the three fertilizers are included in the same experiment, 32 plots are required for 4 replications. But of these, 16 receive nitrogen, and 16 are without nitrogen. Thus there is 16-fold replication on the average response to nitrogen, and similarly for potash and phosphate. To obtain equal replications in three separate experiments, we would require 96 plots. Further, in the combined experiment, information is obtained on any interactions between the effects of the three fertilizers. In fact, if interactions are to be studied, factorial design is necessary. The arguments in favor of factorial design will be found in more detail in the references.

A consequence of the use of factorial design is that the number of treatments tends to become large, or at least sufficiently large so that the simple Latin square requires too many replications and the randomized block design may not be very efficient. The block size may be reduced by giving up the rule that all treatments must appear in the same block. If this is done, the differences between block totals also represent some of the treatment differences, which are said to be mixed up, or "confounded" with blocks. In constructing these designs, the object is to confound only those treatment differences in which we are not particularly interested. The average effects of any factor, and the interactions between pairs of factors (or first-order interactions, as they are called) must be kept clear of blocks if possible, since these usually form the main object of study. Experience has shown, however, that in many types of field experiment, interactions of the second or higher orders are nearly always small, and in constructing confounded designs we try to confine the confounding to these high-order interactions.

In analysing the results of a confounded experiment, the degrees of freedom between blocks are taken out as usual, i.e., the error is derived entirely from within-block comparisons. It is important to ensure that any treatment effects which are compared with this error are themselves derived from within-block comparisons. In the text-books on confounding, the degrees of freedom which are confounded are clearly stated, so that this usually presents no difficulty. A simple test, in doubtful cases, is to calculate the degree of freedom in question from the results, and either (1) see whether the numerical result would be changed by adding a constant amount, say 50, to all the plots in any one block, or (2) verify that in every block the treatment comparison contains an equal number of positive and negative signs. An example of the use of this test will be given below.

A treatment comparison which is confounded may either be completely mixed up with block effects, or only partially so. In the former case, no within-block estimate of the treatment effect is possible. In the latter, a within-block estimate can be made and compared with the ordinary error, though it will not be derived from as many replications as treatment effects which are unconfounded. These two cases may be illustrated from the $2 \times 2 \times 2$ design. If the eight treatments are o, a, b, c, ab, ac, bc, abc, the second-order interaction is written:

$$abc + a + b + c - ab - ac - bc - o$$

Thus the block size may be reduced from eight to four by putting abc, a, b, and c, in the one block, and ab, ac, bc, and o in the other. If four replications of this type are run, the second-order interaction is completely confounded with blocks. Clearly no within-block estimate of this quantity can be made, since all plots in the same block contribute either a + sign or a - sign to the estimate. The four replications might, however, have been laid out as shown on the following page.

Replication

1		2		3		4	
a	ab	o	a	o	a	o	b
b	ac	c	ac	b	ab	a	ab
c	bc	ab	b	ac	c	bc	c
abc	o	abc	bc	abc	bc	abc	ac
ABC		AB		AC		BC	

In this arrangement, the second-order interaction ABC is confounded in the first replication, AB in the second replication, AC in the third, and BC in the fourth. In obtaining within-block estimates of these effects, no use can be made of the particular replication in which each is confounded, but within-block estimates can be secured without any adjustment from each of the three other replications. Thus the design provides three replications on each of these interactions, and four replications on all main effects. In the popular sense, we might say that each of the interactions was one-quarter confounded, or that the relative information on the interactions was three-quarters.

The following example is intended as an exercise in finding out what is confounded in a particular design. For simplicity, the treatments are arranged systematically in the blocks.

Blocks

1	2	3	4
s_1	s_4	s_3	s_2
$npks_1$	$npks_4$	$npks_3$	$npks_2$
ns_2	ns_3	ns_4	ns_1
pks_2	pks_3	pks_4	pks_1
ps_3	ps_2	ps_1	ps_4
nks_3	nks_2	nks_1	nks_4
ks_4	ks_1	ks_2	ks_3
nps_4	nps_1	nps_2	nps_3

This design is clearly of the $2 \times 2 \times 2 \times 4$ type, involving 32 treatments. Since there are four blocks of eight plots each, three treatment degrees of freedom are confounded with blocks. The main effects of the s-factor are clear of block effects, since each block contains two plots at each level of s. Also, each block contains all the eight possible combinations of the n,p,k, factors, so that all main effects and interactions between these factors are clear. Hence the confounding must be confined to the interactions between the s-factor and the n,p,k, factors. Consider the NS interaction. This may be found by calculating the response to n at each of the four levels of s, and comparing these responses. It will be found on inspection that these responses are also clear of blocks, since in any block the pair of plots at a given level of s consists of one without n and one with n. Hence the NS interaction is clear of blocks, and the same is true of the PS and KS interactions.

For those who wish to carry the example further, it may be verified that

Block 1 + Block 2 - Block 3 - Block 4 is the NP ($s_1 + s_4 - s_2 - s_3$) interaction,

Block 1 - Block 2 + Block 3 - Block 4 is the NK ($s_1 + s_3 - s_2 - s_4$) interaction,

Block 1 - Block 2 - Block 3 + Block 4 is the PK ($s_1 + s_2 - s_3 - s_4$) interaction.

Thus, NPS, NKS, and PKS are each one-third confounded. The third-order interaction NPKS is clear.

The table 1 on the following page gives a summary of the principal designs involving confounding which have proved of common utility. Where all factors have the same number of levels, such as 2, 3, or 4, the designs are easy to construct and analyse, and it is usually possible to avoid confounding any two-factor interactions. If the factors have different numbers of levels, it is more difficult to confine the confounding to the higher-order interactions, and the analysis is also more involved.

The principal object of the table is to indicate what interactions must be sacrificed, at least partly, to obtain a given reduction of block size. The experimenter who wishes to use these designs, without learning the details of their construction, should satisfy himself that they do not seriously confound any interaction which is of particular interest. In the 2^n systems, there is considerable freedom of choice in the interactions which are to be confounded. With other types, there is much less choice.

This table refers only to a single replication. With several replicates, these may be made all alike, in which case the confounding is restricted to a few degrees of freedom, though these are rather heavily confounded. Alternatively, the second and further replicates may be constructed with a view to spreading the confounding as evenly as possible among all high-order interactions, so that some information is available on all of these. Designs in which all interactions of a given order are equally confounded are called balanced. Balanced designs have some attractive features: they are more easily analysed than unbalanced designs, some information is obtained on all interactions, and the loss of replication on the partially confounded interactions is reduced to a minimum. Table 2 gives a summary of the balanced designs which can be derived for the factorial systems in table 1, excluding those which require too many replications.

Table 1. Factorial Designs Involving Confounding

Factors ABC...	No. of Treat- ments	Size of Block	Interactions Confounded in a Single Replication
z^3	8	4	ABC
$\oplus z^4$	16	8	ABCD
z^4	16	4	AB, ACD, BCD, or AB, CD, ABCD
$\oplus z^5$	32	8	ABC, ADE, BCDE
z^5	32	4	BD, CE, ABC, ADE, ACD, ABE, BCDE
z^6	64	16	ABCD, ABEF, CDEF
$\oplus z^6$	64	8	ACE, BDE, BCF, ADF, ABCD, ABEF, CDEF
$\oplus 3^3$	27	9	ABC (p)
$\oplus 3^4$	81	9	ABC (p), ABD (p), ACD (p), ECD (p)
$+ 3x2^2$	12	6	BC (p), ABC (p)
$+ 3x2^3$	24	12	BCD (p), ABCD (p)
$+ 3^2 x 2$	18	6	AB (p), ABC (p)
4^2	16	4	AB (p)
4^3	64	16	ABC (p)
$4x2^2$	16	8	ABC (p)
$4x2^3$	32	8	ABC (p), ABD (p), ACD (p)
$+ 4x3x2$	24	12	AC (p), ABC (p)

\oplus Good Latin square designs are also available for these experiments.
(See below.)

$+$ In these cases, only the balanced designs should be used. (See below.)
The symbol (p) denotes that the degrees of freedom are only partially confounded. The factors ABC... are to be read from the left: thus the AC interaction in the 4x3x2 design is the interaction between the factor at four levels and the factor at two levels.

Table 2. Balanced Factorial Designs

Factors ABC...	No. of Treat- ments	Size of Block	No. of Repli- cations	Interactions confounded 1 = 1st order, 2 = 2nd order, etc.
2^3	8	4	4	1&2, $(3/4)$
2^4	16	8	5	2&3, $(4/5)$
2^4	16	4	6	1 $(5/6)$, 2 $(1/2)$
2^4	16	4	4	1 $(5/4)$, 2 $(3/4)$, 3 $(1/2)$
2^5	32	8	5	2 $(4/5)$, 3 $(4/5)$
2^6	64	16	5	3 $(4/5)$, 4 $(4/5)$
3^3	27	9	4	2 $(3/4)$
3^4	81	9	4	2 $(3/4)$
3×2^2	12	6	3	BC $(8/9)$, ABC $(5/9)$
3×2^3	24	12	3	BCD $(8/9)$, ABCD $(5/9)$
+ $3^2 \times 2$	18	6	3	AB $(7/8)$, ABC $(5/8)$
$3^2 \times 2$	18	6	6	AE $(7/8)$, ABC $(5/8)$
4^2	16	4	3	1 $(2/3)$
4^3	64	16	3	2 $(2/3)$
4×2^2	16	8	3	2 $(2/3)$
4×2^3	32	8	3	2 $(2/3)$
$4 \times 3 \times 2$	24	12	3	AC $(8/9)$, ABC $(5/9)$.

1/. The figures in brackets indicate the amount of available information on the partially confounded interactions, relative to that on the unconfounded effects. Thus, in the second 2^4 design in blocks of 4 plots, the first-order interactions are estimated from three replications, as against four for the main effects.

+ This design is not completely balanced, but is fairly easy to analyse and requires only three replicates.

Even if the first-order interactions are required with full accuracy, it may be better to use a design, such as the 3×2^2 , which confounds these slightly, than to use randomized blocks, for if the reduction of block size in this case decreases the error to less than $8/9$, both main effects and first-order interactions will be more accurately determined than in randomized blocks. Experience will show whether the gain is usually of this magnitude.

Confounding in Latin squares. Just as the block size may be reduced by making certain treatment comparisons correspond to the differences between blocks, some factorial designs may be put in Latin squares so that certain treatment differences are confounded with the rows and columns of the square. The possibilities are much more limited than with randomized blocks, since the design has to be arranged so that high-order interactions are confounded both with rows and columns. However, some designs have been found which are likely to be useful, and may compare favorably with the corresponding randomized block designs if the required number of replications happens to be suitable. The most promising are as follows:

2^4 in an 8×8 Latin square (4 replications)

2^5 in an 8×8 Latin square (2 replications)

2^6 in an 8×8 Latin square (1 replication)

3^3 in a 9×9 Latin square (3 replications)

3^4 in a 9×9 Latin square (1 replication)

These designs do not confound any first-order interactions. There is no suitable 3×2^2 design in a 6×6 Latin square, but the $3^2 \times 2$ design may be put in a 6×6 Latin square which retains $3/4$ of the relative information on the 3×3 first-order interaction, and $1/4$ of the relative information on the second-order interaction. Examples of these designs are given below. For practical use, the rows and columns must be randomized. Thus where the square contains more than one replicate, it is not possible to keep the replications separate. A careful study of the designs which follows is a useful exercise for those who wish to become familiar with their construction. The degrees of freedom confounded with rows and columns should be checked.

While these designs are Latin squares in the sense that differences between rows and columns are eliminated from the true error, they are not Latin squares in the original sense that each treatment occurs once in each row and in each column. For this reason, they have been called Quasi-Latin squares.

Useful Latin square designs involving confounding (for practical use after randomizing rows and columns).^{3/}

Notation: 2^n system. The factors are denoted by A, B, C ... The letters a, b, c ... refer to the treatments on the plots. For instance, if the experiment involves all combinations of: no nitrogen and nitrogen (n), two spacings (s_1, s_2) and two sowing dates (d_1, d_2), abc could be taken to represent ns_2d_2 . In this case b denotes no nitrogen, s_2, d_1 .

2^4 design in an 8 x 8 Latin square (4 replications)

There are two alternatives.

Table 3

o ab	a b :	bc ac	abc c	bd ad	abd d	cd abcd	acd bcd	BCD
bc ac	abc c	bd ad	abd d	cd abcd	acd bcd	o ab	a b	BCD
bd ad	abd d	cd abcd	acd bcd	o ab	a b	bc ac	abc c	BCD
cd abcd	acd bcd	o ab	a b	bc ac	abc c	bd ad	abd d	BCD
ABCD		ABCD		ABCD		ABCD		

In this, the third-order interaction and one second-order are completely confounded.

Table 4

c abd	abcd o	b bcd	ad bc	a acd	bd ac	abc d	cd ab	ABC
d bcd	bc ad	a acd	abcd bd	b abc	cd ab	abd c	ac o	ABD
a abc	bd ac	c abd	ab cd	d bcd	abcd o	acd b	bc ad	ACD
b acd	ab cd	d abc	ac o	c abd	ad bc	bcd a	abcd bd	BCD
ABCD		ABCD		ABCD		ABCD		

This retains 3/4 of the information on all second-order interactions and completely confounds the third-order.

3/. I wish to thank Miss Gertrude M. Cox, Iowa State College, for supplying me with copies of these designs.

2^5 in an 8 x 8 Latin square (2 replications)

15

Table 5

o	abe	bc	ace	abd	acd	bcde	de	ABC
bce	ac	e	ab	bcd	d	abde	acde	ADE
cde	abcd	bde	ad	abce	ae	b	c	BCDE
bd	ade	cd	abcde	be	ce	abc	a	
abc	ce	acde	bcd	o	bde	ad	abe	ABD
acd	bcde	abce	c	ade	ab	e	bd	BCE
abde	d	a	be	cde	bc	ace	abcd	ACDE
ae	b	abd	de	ac	abcde	cd	bce	

ACE, BCD, ABDE

ACD, BDE, ABCE

One-half of the information is sacrificed on eight of the ten second-order interactions. (The two which are free would naturally be selected as those in which there was some special interest.) One-half of the information is also lost on four of the five third-order interactions.

2^6 in an 8 x 8 Latin square (1 replication)

Table 6

abcdef	cef	bf	bde	ae	abc	adf	cd	ABC
cde	abce	a	adef	bef	cf	bd	abcdf	ADF
bdf	af	abcef	abcd	c	be	cdef	ade	BDE
acf	bcd	def	o	abd	acde	abef	bce	CEF
bc	acd	abde	abf	df	bcdef	e	acef	ACDE
ef	abdef	acdf	ace	bcde	d	bcf	ab	BCDF
ad	b	ce	cdf	abcf	aef	abcde	bdef	ABEF
abe	de	bcd	bcef	acdef	abdf	ac	f	

ABE, BDF, CDE, ACF, BCEF, ABCD, ADEF

Eight of the twenty second-order interactions and six of the fifteen third-order interactions are completely confounded. The unconfounded third- and higher-order interactions gives 16 degrees of freedom which may be used for an estimate of error.

3^n system: Notation. Here it is more convenient to denote the three levels of any factor by 0, 1, 2. Thus the treatment 021 means the zero level of the first factor, the highest level of the second and the middle level of the third.

3^3 in 9 x 9 Latin square (3 replications)

Table 7

000	101	202	011	112	210	022	120	221
101	202	000	112	210	011	221	022	120
202	000	101	210	011	112	120	221	022
012	211	110	222	121	020	200	001	102
110	012	211	020	223	121	001	102	200
211	110	012	121	020	222	102	200	001
021	220	122	002	201	100	010	212	111
122	021	220	100	002	201	212	111	010
220	122	021	201	100	002	111	010	212

Four of the eight degrees of freedom for the second-order interaction are completely confounded.

3^4 in a 9 x 9 Latin square (1 replication)

Table 8

0000	1022	2011	0112	1101	2120	0221	1210	2202
1012	2001	0020	1121	2110	0102	1200	2222	0211
2021	0010	1002	2100	0122	1111	2212	0201	1220
0111	1100	2122	0220	1212	2201	0002	1021	2010
1120	2112	0101	1202	2221	0210	1011	2000	0022
2102	0121	1110	2211	0200	1222	2020	0012	1001
0222	1211	2200	0001	1020	2012	0110	1102	2121
1201	2220	0212	1010	2002	0021	1122	2111	0100
2210	0202	1221	2022	0011	1000	2101	0120	1112

Half of the degrees of freedom for each of the four second-order interactions are completely confounded.

3 x 3 x 2 in a 6 x 6 Latin square (2 replications)

Table 9

100	020	210	011	201	121
010	200	120	221	111	001
220	110	000	101	021	211
021	211	101	200	120	010
201	121	011	110	000	220
111	001	221	020	210	100

The relative information on the first-order interaction of the 3 x 3 factor is 3/4. Only 1/4 of the relative information is obtained on the second-order interaction, so that these four degrees of freedom are probably best put in with the error.

Varietal trials

In plant selection and breeding work, it is frequently necessary to test in the same experiment a large number of varieties. The problem of constructing accurate designs for these trials has received considerable attention in the past 20 years, and several types of solution have been produced. As in factorial design, the difficulty arises because of the large number of treatments. The device of confounding unimportant treatment comparisons can seldom be used, however, because as a rule all comparisons between pairs of varieties are required with equal precision, except possibly comparisons between new varieties and a control or standard variety, on which higher precision may be desired. The principal designs which have been suggested are as follows:

Randomized blocks. As pointed out before, this may be inaccurate because of the large number of treatments in the block. However, in the earlier stages of breeding or selection, where the number of varieties is high, the plots are often kept very small, and it may be possible to select uniform sites. No statistical analysis may be wanted at this stage, but if one is required for some of the more promising varieties, the rejects may be omitted without any complications. There is also complete freedom in the number of replicates.

Systematic controls. This method is an attempt to measure the fertility variations within the blocks, and to use the measure to increase accuracy. Plots of a control variety are placed systematically throughout the block, and the experimental varieties are randomized in the remaining plots. For example, a block might be planted as follows:

c₁ v₁ v₂ v₃ c₂ v₄ v₅ v₆ c₃ v₇ v₈ v₉ c₄ v₁₀ ...

c₈ v₁₉ v₂₀ v₂₁ c₉ v₂₂ v₂₃ v₂₄ c₁₀ v₂₅ v₂₆ v₂₇ c₁₁ v₂₈ ...

in which c₁, c₂, ... are all the same variety.

From the yields of the controls, indices are constructed of the fertility levels of the plots carrying the trial varieties. There are several ways of doing this. For instance, the index for v_1 might be taken as $(3/4)c_1 + (1/4)c_2$, or the c 's in the second row might be brought into the formula. When the indices have been constructed, there are also several ways of using them. One is to subtract each index from the yield of the variety, and use the resulting figure as a corrected or improved estimate of yield. This method, however, puts too much faith in the index, and will in fact make the yields less precise if the index is a poor one. The correct method is to estimate the regression of the actual on the indicated yield by what is now known as an analysis of covariance. The regression coefficient gives a factor by which the indices are multiplied before subtracting them from the actual yields. If the controls are poor indicators, the factor is small. If the controls are good indicators, a substantial reduction in the standard error may result.

The use of systematic controls is by no means certain to be more accurate than randomized blocks, because it requires more land for a given amount of replication. Suppose there are 100 experimental varieties, in threefold replication. With the distribution of controls shown above, there is one control to every three varieties, and the experiment requires about 400 plots. In randomized blocks 4 replications could be grown on the same piece of land. Thus the introduction of controls must reduce the variance to less than $3/4$ before there is any gain in precision. From experiments with systematic controls, and from uniformity trial data, we could study whether the gain is likely, on the average, to be of this magnitude. In my own opinion, it is doubtful whether the method is likely to be a substantial improvement on randomized blocks, though it might be useful if plenty of land was available, but the supply of experimental seed limited for varieties other than the control.

Ingenious attempts have been made by Richey and later by Papadakis to make the varieties serve as their own fertility indices. The experiment is planted in randomized blocks, and fertility indices are constructed from the differences between the yield of each plot and the mean yield of the variety grown on the plot. These methods do not lead to an exact test of significance, because of the intercorrelations which they introduce, but they may be made approximately valid, provided that there is sufficient replication. They might be useful for a large experiment in randomized blocks in which the variation within blocks turned out to be large.

Random controls. This method aims directly at reducing the size of block. Suppose that there are 100 varieties, one being a control. The 99 varieties, other than the control, may be divided into 11 groups of nine each. Each group, with the control variety, is put in a randomized blocks experiment with 10 plots per block. Thus, if three replications were required, the experiment would require $3 \times 10 \times 11 = 330$ plots, as against 300 in randomized blocks.

By reducing the size of block from 100 to 10, we expect the error mean square to be reduced. To obtain some idea whether the design is likely to be promising, we may calculate roughly by how much the variance must be reduced to make the design superior to randomized blocks. We shall take, as our measure of accuracy, the average error variance of the difference between two varieties. For two varieties which occur in the same group, this is simply $2\sigma^2/r$, where r is the number of replications (in this case three). Two varieties not in the same group must be compared through their controls, i.e. by calculating $(v_1 - c_1) - (v_2 - c_2)$, where c_1, c_2 , are the controls which go with v_1, v_2 respectively. The variance of this comparison is $4\sigma^2/r$. The total number of comparisons between pairs of varieties is $100 \times 99/2 = 4950$. Of these, $11 \times 10 \times 9/2 = 465$ are within-group comparisons, having variance $2\sigma^2/r$. Thus the average variance is

$$2\sigma^2/r(465+2 \times 4485)/4950 = 2(1.906)\sigma^2/r$$

Hence, the variance must be reduced to $1/1.906 = 0.525$ of its value for blocks of 100 plots, before this design becomes more accurate than randomized blocks. If the extra land required is taken into account, this figure is further reduced in the ratio 300/330, i.e. to 0.48. It is very unlikely that the decrease in block size would have such a large effect, and this design as it stands is not hopeful.

Many variants of the method are, however, possible. For instance, the varieties could be divided into groups of eight, and two control plots put with each group in randomized blocks of 10 plots. The effect is to increase the accuracy of comparisons between varieties not in the same group, at the expense of using slightly more land. The net result is to increase the "efficiency factor" from 0.48 to 0.57. This is rather more promising, but if the device is carried further, three controls being put in each group, the efficiency factor is 0.55, and with more controls it begins to fall again. With two and three controls per group, the accuracy may be increased slightly by choosing these as different varieties, instead of all the same variety.

A more important change is to attempt to use the intergroup information directly. With one control, say, the eleven groups may themselves be regarded as eleven treatments, which are being compared in threefold replication on plots 10 times the size of the original plots. If the blocks of size 10 are themselves arranged in a randomized blocks design, the whole experiment becomes a split-plot design, in which the main-plot treatments are the comparisons between groups, and the sub-plot treatments the comparisons within groups. In this case, there is no need to carry the extra controls, as comparisons between groups may be made directly by means of the whole-plot error. Suppose the varieties are split into 10 groups of 10 varieties each, with threefold replication. The analysis of variance runs as follows:

	Degrees of freedom
Replications	2
Between groups	9
Error	18
Within groups	90
Error	180

The variance of the difference between two varieties in the same group is, of course, derived directly from the within-groups (or sub-plot) error. The error variance of the difference between two varieties not in the same group is composed partly of the sub-plot and partly of the main-plot error. If there are g varieties in each group, the general formula for the variance is: $\frac{2}{r} \frac{(S + M - S)}{g}$, where M, S are the main- and sub-plot error mean squares.

For instance, if the main and sub-plot error mean squares in the above experiment were 80 and 20 respectively, the error variance of the difference between two varietal means would be $2 \times 20/3 = 13.3$, for varieties in the same group, and $2(20 + \frac{80-20}{3})/3 = 17.3$ for varieties in different groups.

10

This design is the most attractive of the set which we have just been considering; it avoids the duplication of extra controls and requires no more land than randomized blocks. Its chief disadvantage is the relatively lower accuracy on comparisons between groups, which constitutes the majority of the comparisons between pairs of varieties. It might be a suitable design if the varieties were grouped genetically into families, comparisons within families being of more interest than those between members of different families. If, however, the experimenter wishes all comparisons between pairs of varieties to be of equal precision, the split-plot design can be improved upon. This is easily seen if we bear in mind that varieties which are put in the same group are more accurately compared than varieties which are put in different groups. The split-plot design keeps the same varieties together in all replications, thus accentuating the differences in precision. If equal accuracy is aimed at, it is clearly better to make the opposite rule, that varieties which have appeared together in the first replication must not appear together in further replications. Designs with this property constitute the fourth--and most recent--solution to this problem.

Lattice (or quasi-factorial) designs. Consider the construction of this type of design for our example of 100 varieties in three replications. The first replication is easy to construct--any division of the 100 varieties into 10 groups of 10 will do. Suppose that these are as follows:

Blocks			
1	2	10
v_1	v_{11}	v_{91}
v_2	v_{12}	v_{92}
v_3	v_{13}	v_{93}
v_4	v_{14}	v_{94}
v_5	v_{15}	v_{95}
v_6	v_{16}	v_{96}
v_7	v_{17}	v_{97}
v_8	v_{18}	v_{98}
v_9	v_{19}	v_{99}
v_{10}	v_{20}	v_{100}

The second replication may also be constructed to satisfy our rule, by putting together all varieties in the same row in the above table. Thus v_1 appears with v_2 , v_3 , v_4 , v_5 , v_6 , v_7 , v_8 , v_9 , and v_{10} in the first replication, and with v_{11} , v_{21} , v_{31} , v_{41} , v_{51} , v_{61} , v_{71} , v_{81} , and v_{91} in the second replication. The second replication reads as follows:

Blocks			
1	2	10
v_1	v_2	v_{10}
v_{11}	v_{12}	v_{20}
v_{21}	v_{22}	v_{30}
v_{31}	v_{32}	v_{40}
v_{41}	v_{42}	v_{50}
v_{51}	v_{52}	v_{60}
v_{61}	v_{62}	v_{70}
v_{71}	v_{72}	v_{80}
v_{81}	v_{82}	v_{90}
v_{91}	v_{92}	v_{100}

The third replication requires a little more thought. No two varieties which are in the same row or in the same column in replication 1 must occur together in a block, since any such pairs have already appeared together either in replication 1 or in replication 2. If any Latin square design is superimposed on replication 1, and the varieties which have the same Latin letter are put in the same block in replication 3, the conditions are satisfied. Similarly, the construction of a fourth replicate requires a Graeco-Latin square. No one has succeeded in constructing a 10x10 Graeco-Latin square, and mathematicians are of the opinion that none exists. Such designs can, however, be constructed for all sizes of block up to 20, except 6 and possibly 10, 14, and 18.

As with the designs previously considered, the next step is to calculate the "efficiency factor" of this design relative to 3 randomized blocks of 100 plots. Actually, this design has the important property that it can be analysed as a randomized blocks experiment of 100 varieties in three replicates, the error mean square, as calculated from the analysis of variance, being an unbiased estimate of the true average error. Thus, this design cannot be less accurate than randomized blocks, for if the reduction of block size from 100 to 10 brings a small or negligible decrease in the standard error, the experiment may be analysed as a 100 x 3 randomized blocks design. This is an important advantage which this design holds over the types previously considered.

The details of the within-blocks analysis will be found in the references, but some idea of the general method will be given here. In performing the analysis of variance, the 2 degrees of freedom for replicates and the 27 degrees of freedom for blocks in the same replication are found as usual.^{4/} The varieties sum of squares must, however, be adjusted so that it is composed entirely of comparisons within the blocks of 10 plots. The easiest way to do this is to divide the 99 degrees of freedom between varieties into 9 between the columns of replication 1, 9 between the rows of replication 1, 9 between Latin letters and the remaining 72. The first 9 are completely confounded with blocks in replication 1, but are clear in the other two replicates. They may therefore be calculated without any difficulty from these replicates. Similarly the 9 degrees of freedom for rows are calculated from replicates 1 and 3, and the Latin letters from 1 and 2. The remaining 72 degrees of freedom are clear of blocks in all three replicates.

The reduction in block size from 100 to 10 has thus been gained at the expense of reducing the number of replications from 3 to 2 of the 99 degrees of freedom between varieties. These 27 degrees of freedom have $3/2$ times the error variance which they would have had in threefold replication. The average error variance for the 99 degrees of freedom is $(27 \times 3/2 + 72)/99 = 25/22$, relative to 1 for threefold replication on all 99 degrees of freedom. Thus the reduction in block size must decrease the error variance to $22/25 = 0.88$, or less, if the within-blocks analysis is to be more accurate than the randomized blocks analysis. The general formula for this efficiency factor is $(p + 1)/(p + 2 - 1/2)$ for p^2 varieties in blocks of size p , (3 replicates).

^{4/}. At the present time the case with three replicates is being considered.

In a recent article, Yates has shown how to combine the randomized blocks and the within-blocks analysis so as to obtain the most accurate comparison between varieties which the experiment provides. The computational details have not yet appeared in print for all types of design, but they have been worked out and should appear shortly. This analysis becomes the randomized blocks analysis if the reduction in block size brings no reduction in error; it becomes the within-blocks analysis when the reduction in block size brings a great decrease in the error. In the intermediate range, which probably contains most of the results likely to be obtained in practice, it is more accurate than either of those. The analysis is somewhat more laborious than the randomized blocks analysis, though not excessively so. The randomized blocks analysis may, however, be used for data which are not much affected by soil fertility variations, or for subsidiary measurements on which it is not considered worth while spending the extra time.

The above design does not give exactly equal accuracy on all varietal comparisons, since some pairs of varieties never appear together in the same block. The discrepancy is, however, small. In the above example, the error variance of the difference between two varietal means is $22\sigma^2/30$, for varieties which appear in the same block, and $23\sigma^2/30$ for those pairs which do not, σ^2 being the error mean square. This result was obtained from the within-blocks analysis; the difference is still smaller with the new method of computation.

Types of Lattice design. In the above designs, the number of varieties must be a perfect square. The most useful range for the designs is shown below:

Number of varieties	25	36	49	64	81	100	121	144	169	196
Size of block	5	6	7	8	9	10	11	12	13	14

For larger number of varieties, the block size tends to become large. Any number of replicates may be used, though some care is needed in constructing the appropriate design. For the 6×6 , 10×10 , and 14×14 designs, the condition that no two varieties shall occur more than once in the same block cannot be carried beyond three replicates (as shown above). However, in these cases four replicates can be obtained by duplicating the design in two replicates, and similarly six replicates from the three-replicate design. Five are less convenient, but can be arranged.

To avoid the restriction that the number of varieties must be a perfect square, Yates also introduced designs for other numbers in two and three replicates. These are, however, more troublesome to analyse, while the restriction does not seem to have troubled the plant breeders, so far as the designs have been applied in practice.

Balanced designs. Most of these designs can be arranged so that all varietal comparisons are of equal accuracy, though the necessary number of replicates rapidly becomes large. In the 5×5 design, for example, each variety occurs with four others in a single block, and has 24 possible companions in all. Hence six replicates are necessary. No such design is possible for 36, 100, or 196 varieties and none has yet been found for 144. For the other numbers, the number of replicates required for complete balance is as follows:

Number of varieties	25	49	64	81	121	169
Number of replications	6	8	9	10	12	14

These designs are constructed by extending the method described for the 10×10 square, until every variety has appeared once in a block with every other variety. In addition to giving all comparisons with equal precision, they are very simple to analyse, and are to be preferred if the amount of replication is not excessive.

Balanced designs in Latin squares. The possibility of arranging these designs in Latin squares has also been investigated. In this case, varietal comparisons are confounded both with rows and columns of the squares. This doubles the number of degrees of freedom confounded in each replication. For the squares of odd side, it enables a balanced design to be constructed in half the number of replicates. A completely balanced 5×5 design may be put in three replicates, a 7×7 in four, a 9×9 in five, and an 11×11 in six. The 8×8 design, however, still requires nine replicates, since nine is an odd number. The possibility of obtaining complete balance, with relatively few replicates, by the use of Latin square designs should be borne in mind. In this respect, the usual roles of the Latin square and randomized blocks are reversed.

These designs also possess the property that they can be analysed as randomized blocks of p^2 plots, and cannot be less efficient than the latter.

The cubic Lattice. The above designs may be expected to deal with numbers of varieties up to 200. Beyond that, even the reduced block is becoming rather large. To obtain a more severe reduction in block size, Yates constructed a design for p^3 varieties in blocks of size p . The principles of construction are similar to those for the above designs, different varietal comparisons being confounded in successive replications. The design requires three replicates, and can also be planned in six replicates by duplication. The relation between number of varieties and block size is shown below.

Number of varieties	125	216	343	512	729	1000
Size of block	5	6	7	8	9	10

The gaps between the admissible numbers of varieties are unfortunately large. It is difficult to say whether this design will prove more efficient than the p^2 designs in the range between 100 and 216 varieties. Much will depend on the heterogeneity of the site. For numbers greater than 216, it seems likely to be the most suitable design, though practical experience will show.

Nomenclature. These designs were at first called pseudo-factorial, because certain parts of their statistical analysis are best understood by pretending that they are factorial designs. Later, the name was changed to quasi-factorial, this being considered more suitable etymologically than the original hybrid word. These names, and the more detailed names which were used to distinguish the various types, are rather complicated. More recently, Yates has introduced a simpler nomenclature. The p^2 design in blocks of size p is called a Lattice. The balanced designs are called balanced lattices, and the Latin square designs are called Lattice squares. The p^3 design in blocks of size p is the cubic lattice. The new names are easier to remember than the old ones. These remarks are included because the changes in nomenclature may produce confusion in reading the literature.

Before summarizing the above discussion on varietal trials, one further type of design will be described. This was originally produced with a different purpose in mind, but has been found to be fairly appropriate in varietal trials.

Balanced incomplete blocks. In field trials, there is seldom any natural restriction on the number of plots which are put in a block. With other types of experiment, however, the natural size of the block may be completely fixed, or variable only within small limits. For instance, in experiments involving the inoculation of plants with a virus disease, different plants may vary considerably in their susceptibility, and to a lesser extent, different leaves on the same plant. Here the natural block is the two halves of the same leaf, and if the technique permits the separate inoculation of the two halves, very accurate comparisons may be made between the effects of different treatments. If a treatment must be applied to the whole of a leaf, the natural block is the plant, on which only four or five suitable leaves may be available for inoculation. Similarly, in animal experiments where it is important to equalize for litter and sex differences, the number of animals of the same sex available from the same litter may be small.

These considerations cause no difficulty so long as the number of treatments to be compared is no larger than the number of units in the block. With larger numbers of treatments, a demand arises for designs comparing t treatments in blocks of size k , where k is less than t and need not be a factor of t (as it is in the Lattice designs). The principles of construction are the same as in the balanced Lattice designs, i.e., any pair of treatments must occur in the same block an equal number of times. The designs with small numbers of treatments are not difficult to construct. With 7 treatments, a, b, c, d, e, f, and g, in blocks of 3 units, for example, the simplest design runs as follows:

Block

1 abc	2 ade	3 afg	4 bdf	5 beg	6 cef	7 cdg
----------	----------	----------	----------	----------	----------	----------

Each treatment occurs once with any other treatment in the same block. Three replications are necessary. These designs give all treatment comparisons with equal accuracy, and are easy to analyze. They cannot, as a rule, be arranged in separate replications, so that the alternative randomized blocks analysis is not possible. However, as with the Lattice designs, the information in the interblock comparisons may be utilized to improve the within-block analysis.

A table of the available designs is given in reference 18, page 38. Frequently the design requires more than ten replications, which is excessive for small experiments.

Use of balanced incomplete blocks in varietal trials. For certain numbers of varieties between 25 and 100, these designs require no more replication than the corresponding lattice designs, and form a useful addition to the repertoire of balanced designs. The complete set of balanced designs with not more than 10 replications is shown below:

Lattice designs			Incomplete blocks		
No. of varieties	Size of block	No. of replicates	No. of varieties	Size of block	No. of replicates
25	5	6	31	6	6
49	7	8	57	8	8
64	8	9	73	9	9
81	9	10	91	10	10

To these, of course, should be added the lattice square designs.

Discussion of varietal trial designs. No extensive examination has been made of the relative merits of the method of systematic controls and of the lattice designs. The former has the advantage of giving all comparisons with equal accuracy, without any restriction on the number of replicates. On the other hand, it is definitely less accurate than the lattice designs when there is no soil heterogeneity within complete replications. Whether it is likely in any circumstances to be substantially more accurate than the lattice designs depends on the pattern of fertility variation within complete replications. If there were a marked fertility gradient in each of two directions, a suitable arrangement of controls might give a better elimination of the effects of the gradient than a lattice design in blocks. Examination of uniformity trial data would be necessary to assess the average relative accuracy of the two types.

The use of random controls to reduce block size appears definitely inferior to the lattice designs, both in its over-all accuracy and in the wide discrepancy between the relative precision of comparisons in the same block and comparisons in different blocks.

Among the lattice designs, the balanced arrangements, either in Latin squares or randomized blocks, are preferable if the replication required is not too great, since they give equal accuracy on all comparisons and are easy to analyze. The incomplete block designs also possess these properties, and are only slightly less convenient because they cannot be arranged in separate replications, and hence do not permit the randomized block analysis, which may occasionally be useful. If the replication demanded by the preceding designs is too great, the lattice designs in two, three, or four replications may be used.

For numbers of varieties greater than 200, the cubic lattice should be tried. This has the disadvantage that it must be run in a multiple of three replications (otherwise the analysis is complicated). It also requires rather more computation than the other designs, but on the score of accuracy it appears the most promising of these designs to deal with really large numbers of varieties.

Youden squares. One further extension of the balanced incomplete blocks design may be mentioned. This is a rearrangement of the design so that fertility variations in two directions may be taken out. For instance, the design given above for 7 treatments in blocks of 3 units might be arranged as follows:

	Blocks						
	1	2	3	4	5	6	7
Row 1	a	b	c	d	e	f	g
Row 2	b	d	f	e	g	a	c
Row 3	c	f	e	a	b	g	d

The blocks are arranged so that each row contains all the treatments, thus forming the first three rows of a Latin square. Differences between blocks and between rows may both be taken out of the estimate of error. This design is particularly suitable for virus experiments where the unit is a single leaf, because there may be a gradient in susceptibility from the highest leaf of a plant to the lowest. Here the blocks consist of plants, and the rows are the positions of the leaves on a plant. The same principle may apply in other types of experiment in which the incomplete blocks design is used.

Other problems of design. With the great increase in replicated experiments in many branches of research during recent years, it would be rash to prophesy what are likely to be the important problems of design in the future. There are one or two types already appearing which have not been considered here. One is the imposition of a factorial type of design on varietal trials containing a large number of varieties. This may arise either through the addition of width of spacing, time of sowing, or fertilizer comparisons, or because the varieties themselves show some kind of grouping, being for example all crosses with three standard varieties. If a balanced design is used, the subdivision of the treatment degrees of freedom in any way presents no difficulty. With unbalanced designs, a statistician should be consulted before proceeding with the trial, to insure that treatment comparisons which are of interest can be isolated without excessive computational labor.

Another problem concerns experiments in which residual effects enter. In experiments on milk production in cows, the lactation period is sometimes divided into two, three, or four periods, different feeding treatments being given in different periods. This eliminates the large source of error arising from differences in the milk yield of different cows. However, the effects of a treatment may persist into the next period, when the animal is receiving another treatment. At present, this is partially taken care of by arranging the design so that each treatment follows every other treatment an equal number of times, and by exercising care in the analysis. Better methods may be developed.

Many new problems arise, both in design and analysis, in experiments involving a rotation of crops. The object of study may be the effects of different fertilizers on a fixed rotation of crops, or the effects of different crop sequences in maintaining or building up the fertility of the soil. In the latter case, should one or more indicator crops be grown at fixed intervals to assess the effects of the crop sequences, or should the crops themselves be made to form the indices of their performance? Since rotation experiments are costly, little experience has been obtained to assist in answering such questions, but as time goes on, the general principles of good design should become clearer.

Notes on the use and interpretation of "experimental error"

The choice of an estimate of experimental error for use in tests of significance appears to present considerable difficulty to research workers who use the analysis of variance to assist them in the interpretation of their results. Part of the difficulty may arise from the way in which the subject is taught. The student usually begins with designs of the randomized blocks and Latin square type. Here there is only one error term, and that is obtained by subtraction after the easily recognized parts of the total sum of squares have been taken out. The erroneous impression is often produced that in general there is only one error to each experiment, against which everything else may be tested, and that this error is the "random" variation left after taking out everything which can be isolated. These opinions may be remedied by a study of the structure of error terms in the analysis of variance. This lecture might profitably have been devoted to such a study, but instead we want to see how far we are helped simply by common sense ideas of what an error should be.

Two types of test of significance must be distinguished. In the first, which may be called local, we are interested solely in describing the results of a single experiment, with no attempt to predict what would happen if the same treatments were applied under different conditions. In the second, we are trying to make recommendations which will be put to the test under widely varying farming conditions and in seasons different from those in which the critical experiments were made. This raises difficult problems in induction from the known to the unknown, but the problems must be faced whenever a general recommendation is attempted.

Consider first the local tests. A good description of the requirements of an error is given in the following words by Fisher.^{5/} "The very same causes that produce our real error shall also contribute the materials necessary for computing an estimate of it. The logical necessity of this requirement is readily apparent, for, if causes of variation which do not influence our real error are allowed to affect our estimate of it, or equally, if causes of variation affect the real error in such a way as to make no contribution to our estimate, this estimate will be vitiated."

These remarks may seem almost platitudes. But they form a useful criterion for judging a proposed error term. They suggest that attention should first be directed to the treatment differences. From the way in which the experiment was designed and carried out, and from the way in which the treatment differences are calculated, it should be possible to write down at least the principal sources of variation that may affect these differences. The next step is to consider whether the proposed error measures all these sources of variation. Finally, we must insure that the estimated error does not include any sources of error which do not affect the treatment differences. In the standard modern designs, randomization insures that the error, as calculated by the analysis of variance, is an unbiased estimate of the real error of treatment differences. But even in very simple deviations from these designs, an error which seems the obvious choice may fail in both respects. Three examples will be given to illustrate these points.

Example 1. Consider an experiment in which six feeding rations are tested on a group of sixty animals, ten to each treatment. Animals receiving the same treatment are kept in one pen and are fed together, each group receiving ten times the ration for one individual. Treatments are allotted to the groups at random.

The analysis of variance of the results is simple, there being 5 degrees of freedom for differences between treatments, and 54 degrees of freedom for differences within treatments. Does the latter mean square constitute a valid estimate of error for testing the treatments? Following the suggestions in the passage quoted above, we must consider all factors which may influence the real error of the treatment differences, and see whether they are taken into account in the error which is suggested. The difference between the effects of two treatments on, say, the live weight increase, is found by calculating the average increase for one group of ten animals and subtracting the average increase for another group, these groups being kept in different pens. Prominent among the sources of error are the individual variations, from animal to animal, in rate of increase. From statistical theory we know that these variations are estimated without bias in the "within-lots" error, provided that the animals were divided into the ten lots at random. Unless the subdivision was made in this way, there is no guarantee that the within-lots mean square gives a proper estimate of this source of error as it affects the difference between two different groups. If, for instance, the different groups were from different herds, the within-lots term would probably give an underestimate, since it takes no account of a possible consistent superiority of one herd over another. On the other hand, if the groups were chosen so as to be closely alike in their average rate of increase, by putting some sturdy and some weak animals in every group, the within-lots term might give an overestimate. The allocation of treatments to groups at random does not get over this difficulty; it merely insures that we shall not deliberately favor a particular treatment.

Assuming that the animals were allotted to the groups at random, what other sources of variation affect the difference between two groups? Clearly anything which affects all the members of the same pen will do so. If the pens are differently situated as regards temperature, exposure to wind and rain or to strong sunshine, access to water, type of flooring, etc., any of these factors may produce differences in rate of growth between the groups. They are ignored in the proposed estimate of error, which is derived entirely from differences between animals in the same pen. If the pens are near one another and all under uniform conditions, it may perhaps be claimed that the effects of such differences on the increases are negligible, but there is no way of proving that this is so from the results of the experiment.

To turn to the other side of the picture, are there any sources of variation measured in the proposed error, which do not affect the real error? There may be. Bullying of the weaker animals in the pen has frequently been observed in animal experiments. The more sturdy members receive more than their share of the food at the expense of these weaker animals. This has possibly little effect on the average growth rate of the group as a whole, since the same total amount of food is eaten, but it inflates the within-groups error.

These remarks show that the within-groups error leaves a good deal to be desired. The difficulties may be overcome either by replicating the groups, or by resorting to individual feeding.

Example 2. A proposed $3 \times 3 \times 3$ experiment on the bacterial content and acidity of milk runs as follows:

Treatments:

Methods of milking: hand, machine, and combine machine.

Type of cooler: conical, tubular, and can in tank.

Time before cooling: 10, 20 and 30 minutes.

Owing to practical difficulties, the three methods must be tried on different herds. The milk from each herd is divided into nine portions, one to each of the combinations of the last two factors. For each treatment, two separate determinations are made of the acidity and bacterial content. The analysis of variance will run as follows:

	Degrees of freedom
Methods of milking	2
Types of coolers	2
Times before cooling	2
Coolers x times	4
Methods x coolers	4
Methods x times	4
Methods x coolers x times	8
Between duplicates	27

The "between duplicates" term cannot serve as an estimate of error. It measures the sampling error involved in drawing a sample from each treatment to determine the acidity, and also any variations introduced in the purely chemical part of the technique. It does not take into account the differences which may arise between milk kept in one can in one place, and milk kept in another can in a different place. The purpose of taking the duplicates is presumably partly to increase the accuracy, and partly to see whether the errors arising from the chemical determinations form an important part of the total experimental error. If they are important, it might be advisable to increase the number of independent determinations made. If they are negligible, it would probably be sufficient in future to make only one determination on each treatment.

It is clear that there is no proper estimate of error for methods of milking. These are being tested on three different herds, and the experiment provides no estimate of the differences which might exist between these herds if milked by similar methods. It would be useful to have previous records of the performances of the three herds under similar treatment. These would serve as a guide in guessing whether the differences produced by the methods were real.

To test coolers, times before cooling and their interaction, it may be sufficient to think of the experiment as being in three randomized blocks, the herds constituting the blocks. If the nine portions into which the milk from each herd is divided are allotted to the nine combinations of coolers and times before cooling at random, the 16 degrees of freedom for interactions with methods of milking (or herds) give an unbiased estimate of the experimental errors which affect coolers, times, and coolers \times times. However, they also contain any real interaction between these factors and methods of milking. If such interactions are negligible, this need not concern us. Even if the interactions are not negligible, a case might still be made out for using these terms as error, on the grounds that a type of cooler must be superior on all herds if it is to be of general use. If such interactions are suspected, it would, however, be better to plan the experiment so that they can be isolated and studied. This could be done, if practicable, by dividing the milk from each herd into 18 portions, running two replications within each herd. This design would provide an error against which all main effects and interactions (except the main effects of methods of milking) could be tested.

Example 3. This example is taken from a paper^{6/} explaining the method of analysing the results of a long-term experiment in which the treatments remain on the same plots for several years. The experiment consisted of seven randomized blocks of three treatments each. Four harvests are given, at 2-year intervals, so that there are 84 degrees of freedom in all. The experiment should clearly provide a test of the average effects of treatments throughout the four seasons, and of the interactions of treatments with seasons. The author suggests the following analysis:

	Degrees of freedom	Sums of squares	Mean squares
Years	3	1715	572
Blocks	24	959	
Treatments	2	768	384
Treatments x Years	6	323	54
Error	48	3205	67

In this analysis, everything which is easy to separate is taken out, the remainder being put into error. The 48 degrees of freedom for error are actually the pooled errors from the analyses of the individual seasons.

The difference between two treatments is the difference between the mean yield of one set of plots in four seasons, and the mean yield of another such set. This difference is certainly affected by within-years errors. In the analysis of variance, however, each total of four seasons is treated as if it were composed of four independent observations. If there is any positive correlation between the yields of the same plot in successive years, this will not be correct. An error that takes this correlation into account is obtained by adding the four years' results on each of the 21 plots, analysing these figures as a 3×7 randomized blocks experiment. This gives an error term with 12 degrees of freedom. This term contains its proper contribution from within-years variation; it also automatically allows for the effects of any correlation, since it is based entirely upon totals. The 12 degrees of freedom are, of course, a part of the 48 degrees of freedom described as error in the table.

	Degrees of freedom	Sums of squares	Mean squares
Error from totals	12	1674	140
Remainder	36	1531	42
Total	48	3205	67

The error from the totals of the four seasons is more than three times the rest of the error, the difference being highly significant. This confirms that the yields of the same plots were positively correlated in different seasons. It also implies that the use of the pooled error of 48 degrees of freedom results in a serious under-estimation of the true error of the difference between the treatment means.

Which error should be used for testing Treatments x Years? These comparisons are derived from the differences between the results in different years. They are, therefore, not affected by a consistent superiority of one plot over another, and should be tested against the 36 degrees of freedom for the remainder of the error, and not against the error from the totals.

This may be seen in another way. The 12 degrees of freedom error in the above table is actually the interaction of the treatment totals with blocks. Randomization^{7/} guarantees that this error is a proper measure of the true errors affecting treatment totals. Similarly, any effect or interaction in this experiment may be tested against its interaction with blocks. It will be found on examination that the 36 degrees of freedom constituting the remainder of the error are the Treatments x Years x Blocks interaction, and are therefore appropriate for testing Treatments x Years.

Thus the error suggested by the author is less than half the proper mean square for the average effects, and about 50 percent too large for the interactions. This type of data occurs frequently in experimental studies, (e.g., with plots which are sampled or cut several times a year), and the method of analysis deserves some study.

Tests of significance for a group of experiments

Suppose that an experiment has been carried out at a number of centers for the purpose of finding out whether certain treatments may be recommended over the whole of the farming area in which the experiments were situated. As an example, we may consider the results of a series of experiments carried out in England on the response of sugar-beet to the three standard fertilizers, N, P, and K. In one year there were 15 centers, and the analysis for the response to N runs as follows:

	<u>Degrees of freedom</u>	<u>Mean squares</u>
Average response	1	17.195
Response x Centers	14	1.828
Pooled error from individual experiments	246	0.234

These data provide a test of the average response, and of the variation of the response from center to center. Each may be tested against the pooled estimate of error, with 246 degrees of freedom.^{8/} For the average response, this test tells whether the response should be regarded as real over the particular set of centers which were chosen. However, if the object of the experiments is to decide whether nitrogen may safely be recommended for the whole farming area, we must have some guarantee that nitrogen would also give responses if a different set of centers had been chosen. For this purpose, the above test is largely irrelevant, since the "error" term does not take any account of the possibility that the response may vary from center to center. In other words, the above test is a purely "local" one, enabling us to make statements applicable only to the particular set of fields where the experiments were carried out.

- 7/. In this experiment, the treatments were not properly randomized, so that the errors suggested above are somewhat open to suspicion.
- 8/. Assuming that the error variances in the individual experiments were not widely different.

The interaction term in the analysis measures the amount by which the treatment response has varied from center to center (and in addition, of course, the ordinary experimental errors). A test of the average response against the interaction mean square is therefore much more relevant to our purpose. Conclusions should, however, not be based upon the test alone. We must satisfy ourselves that the centers chosen are representative of the average conditions and the range of conditions in ordinary farming. Otherwise the response to nitrogen and its variation at these centers may not be typical of the results which would be obtained in practice. This degree of representativeness could be secured by picking the experimental fields at random from the available fields in the area, but this is not usually practicable, (though I believe it is done in Sweden). As safeguards, we may inquire whether the average yield in the experiments was about the same as the average yield in the whole area and whether a reasonable range of yields was obtained. It is probably also advisable not to force too much uniformity in the farming operations to be carried out at the different centers.

The responses at individual centers should be examined, inspecting particularly any centers where the results were highly divergent. There may be simple explanations of these discrepancies which will cause us to revise a verdict reached simply on the test of significance of average response against interactions. Sometimes one particular type of soil may fail to show responses.

The same considerations apply to a test of treatment effects against the interaction of treatments with years. Here additional caution is needed, because there are seldom more than a few years available, whereas an adequate number of centers can usually be included. It has sometimes been said that any group of years may be regarded as a random selection of all years, because weather variations are random. This is a dangerous assumption for the present purpose, and the experimenter is well advised to consider whether a representative sample of the range of weather and disease conditions was provided in the seasons in which his experiments were carried out.

For instance, in the above series of sugar-beet experiments, the average responses to N, P, and K in the first three years were as follows:

Roots: tons per acre

	Response to		
	N	P	K
1933	+0.64	+0.14	+0.28
1934	+1.07	+0.32	-0.06
1935	+1.12	+0.12	+0.16
Mean	+0.94	+0.19	+0.13

The conclusions would be that a response of just under a ton to the acre may be expected from N, but very small responses from P and K. However, conditions during the growing season were warm and dry in all three years. In 1936, there was a wet season. The responses were:

	N	P	K
	+2.26	+0.84	+0.40

All responses are well outside the range indicated by the first three years. The results in 1937 and 1938 were

	N	P	K
1937	+1.70	+0.55	+0.66
1938	+0.76	+0.39	+0.71

For P and K, the responses in each of the last three years are above any that occurred in the first three. In these experiments, it should be noted, the fields are changed every year, so that there is no question of a cumulative effect. Clearly if the experiments had been terminated at the end of the third year, the expected returns from application of the fertilizers would have been much too small.

It is, of course, not the fault of the statistical methods that these tests begin to fail if the centers or years are unrepresentative. The inductions made from the tests will apply to any set of conditions of which the experiments constitute a random sample. If these conditions are not the conditions in which the experimenter is interested, he should consider whether changes in the planning of the experiments would improve matters. If such changes are beyond his power, added caution is required in making recommendations from the results.

ReferencesRandomized blocks and the Latin square

Several good textbooks are available, e.g.:

1. Fisher, R. A. The design of experiments. Chaps. 4 & 5.
2. Wishart, J. & Sanders, H. G. The principles and practice of field experimentation.
3. Wishart, J. Field trials: their layout and statistical analysis. Imp. Bur. Plant Breeding and Genet.

Methods of calculating the relative efficiency of different designs

4. Yates, F. Complex experiments. Supp. Jour. Roy. Statis. Soc., 1935, pp. 181-247.

The split-plot design

Nos. 3, 4 above. Also:

5. Yates, F. The design and analysis of factorial experiments. Imp. Bur. Soil Science, Tech. Commun. No. 35.
6. Goulden, C. H. Methods of statistical analysis. Chap. 12

Factorial design

Nos. 1, 4, 5 above. For an early account of the general principles, see:

7. Fisher, R. A. The arrangement of field experiments. Jour. Min. Agr. 1926, vol. 33, pp. 503-513.

Varietal trials

A brief discussion of the methods of systematic and random controls is found in:

8. Yates, F. A new method of arranging varietal trials involving a large number of varieties. Jour. Agr. Sci., (England), 1936, vol. 26, pp. 424-455.

For earlier accounts of the method of systematic controls, see:

9. Richey, F. D. Adjusting yields to their regression on a moving average, as a means of correcting for soil heterogeneity. Jour. Agr. Res. 1924, vol. 27, pp. 79-90.
10. Stadler, L. J. Experiments in field plot technique for the preliminary determination of comparative yields in the small grains. Mo. Agr. Exp. Sta. Res. Bul., 1921, No. 49, p. 72.

Lattice designs

A brief account is given in No. 5.

These designs were first described in No. 8. This paper contains the general theory, and worked examples of a 7×7 lattice and a 4^3 cubic lattice. No. 6 contains worked examples of a 5×5 lattice and a 3^3 cubic lattice. A general account, with worked examples of a 5×5 lattice, a 4×4 lattice in three replications, and a 4^3 cubic lattice, is given in:

11. Goulden, C. H. Modern methods for testing a large number of varieties. Dominion of Canada, Tech. Bul. 9, 1937.

A 9^3 cubic lattice is described in:

12. Day, B. B. & Austin, Lloyd. A three-dimensional lattice design for studies in forest genetics. Jour. Agr. Res., 1939, vol. 59, pp. 101-120.

A 3^3 cubic lattice is described in:

13. Dawson, C. D. R. An example of the quasi-factorial design applied to a corn-breeding experiment. Ann. of Eugenics, 1939, vol. 9, pp. 157-173.

The new method of analysis, taking into account both the inter- and intra-block information, is given, for the cubic lattice, in:

14. Yates, F. The recovery of inter-block information in variety trials arranged in three-dimensional lattices. Ann. Eugenics, 1939, vol. 9, pp. 136-156.

Lattice squares

15. Yates, F. A further note on the arrangement of variety trials: quasi-Latin squares. Ann. of Eugenics, 1937, vol. 7, pp. 319-331.

A 5×5 lattice square is analysed in:

16. Weiss, M. G. & Cox, G. M. Balanced incomplete block and lattice square designs for testing yield differences among large numbers of soybean varieties. Iowa State Col. Exp. Sta. Res. Bul. 257, 1939.

Balanced Incomplete Blocks

17. Yates, F. Incomplete randomized blocks. Ann. Eugenics, 1936, vol. 7, pp. 121-140.

A design with 31 varieties in blocks of 6 plots is analyzed in Nos. 11 and 16.

A table of the designs which have so far been discovered and a brief account of the recovery of inter-block information, are given in:

18. Fisher, R. A. & Yates, F. Statistical tables for agricultural, biological and medical research.
19. Yates, F. The gain in efficiency resulting from the use of balanced designs. Sup. Jour. Royal Statis. Soc., 1938, Vol. 5, pp. 70-74.

This discusses an experiment with 4 treatments in blocks of 3 plots.

Youden squares

20. Youden, W. J. Use of incomplete block replications in estimating tobacco-mosaic virus. Boyce Thompson Inst. Contrib., 1937, Vol. 9, pp. 41-48.

Rotation experiments

An introductory account of the problems is given in:

21. Cochran, W. G. Long-term agricultural experiments. Supp. Jour. Roy. Statis. Soc., 1939, Vol. 7, pp. 104-148.